

# Unit I    IDS

- What is Data?  
(Raw facts)
- What is Information? (Useful insights from raw data)
- What is Science? → Knowledge on information which we gain from experiments / Observations.
- What is Data Science? → Developing useful insights from data.  
In order to solve complex problems.
- Why we study Data Science? → To extract knowledgeable insights from data.
- Components of Data Science
  - Data Strategy → [incorporates methods & services]
  - Data Engineering → Understanding data sources
  - Data Modeling & Analysis → Data Pipelining, Enrichment
  - Data Visualization & Operationalization → Data Retention Mang.
  - Graphical representation of data in form of tables, charts & dashboards → Data Science is combination of various statistical models (we analyze the data)

- 
- What is Big Data  
(Refers to huge amount of data that cannot be stored & processed using traditional approaches within a given time frame)
  - Example of Big data (100 MB of file attachment with respect to email can be referred to as Big Data)
  - Big Data Analytics  
(Process of extracting meaningful insights from data)

- 5 V's of Big Data OR Characteristics of Big data
  - Velocity
  - Volume
  - Variety
  - Veracity
  - Value

- Applications of Big Data

- Education
- Healthcare
- Weather forecasting
- Agriculture
- Manufacturing

(Any many more)

- Sources of Big Data

- Social Network Data
- Financial Data
- Multimedia Data
- Mobile Apps Data
- Internet of Things

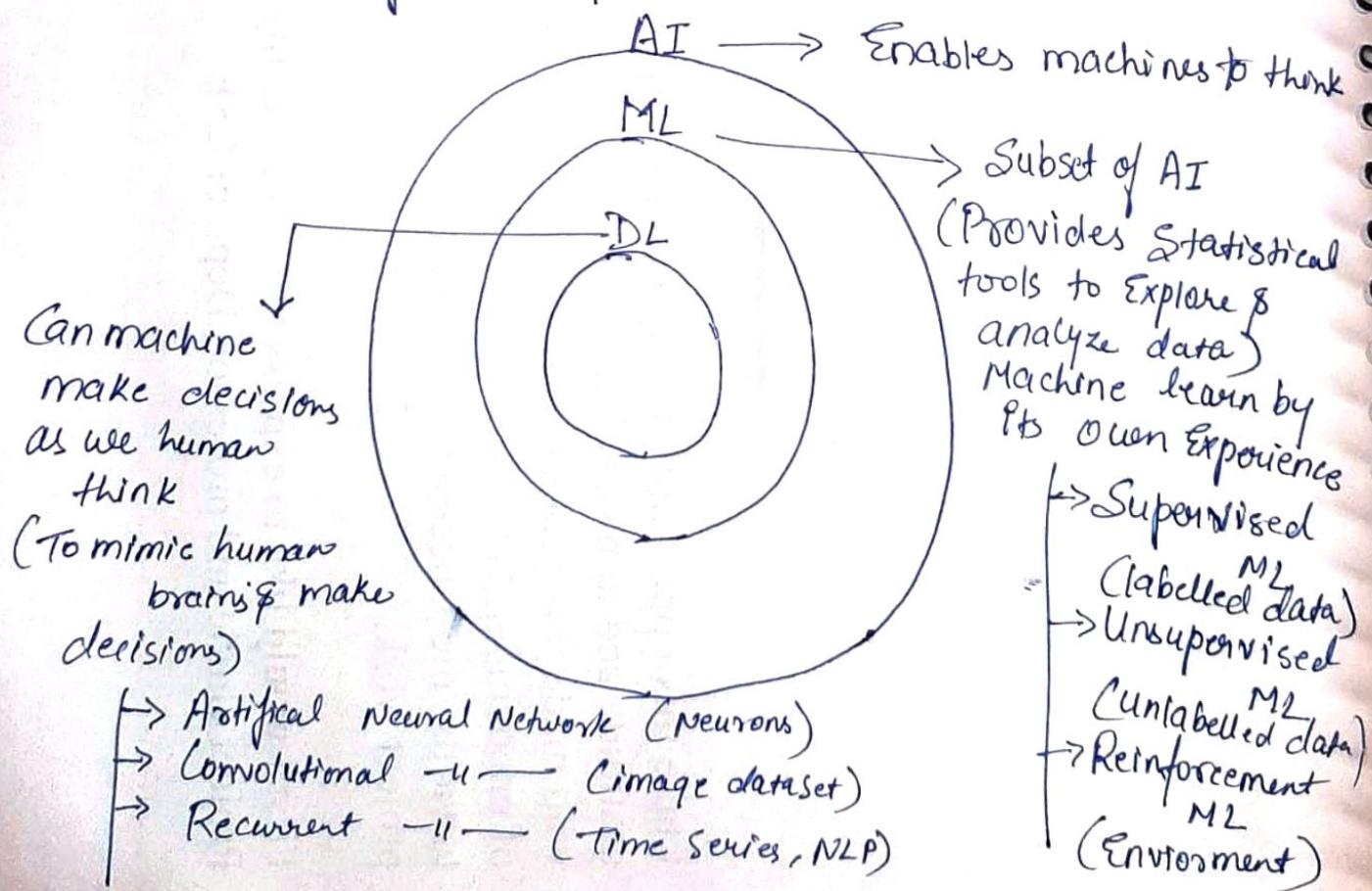
- Big Data Technologies

- Apache Hbase (NoSQL database)
- Spark (Real time processing)
- Talend (S/w Integration Platform)
- Splunk (Log Analysis Platform)
- Hive (Data warehousing)
- Kafka (Messaging System)
- hadoop (Data Storage & Analysis Platform)

- Data Lake (Centralized repository designed to store & process data in secure way, specially large amount of data)
- Big data Challenges
  - Required specialized skills & expertise
  - Can be expensive, resources, S/W
  - Can be challenging to integrate with existing system, Privacy & Security
- Big Data & Data Science
  - Similarities
  - Differences

Difference b/w AI, DS, Deep learning & Machine learning

Artificial Intelligence      Data Science



- Data Analytics vs Data Scientist vs Data Engineer
- Data Analytics (works on Past data & make future predictions)
- Data Scientist (Design & develop new process for modelling the data by using algorithms, Predictive models & makes decisions)
- Data Engineer (Involves in Preparing the data, develop & construct & maintain complex architecture)

### Data Scientist

Skills &  
Job roles

- Statistical & Analytical Skills
- Data Mining
- ML & DL
- deep programming knowledge (R/Python)
- Decision making & softskill

### Data Engineers

Skills &  
Job roles

- Extract ~~load~~ Transform Load (ETL)
- Advanced Programming knowledge
- Hadoop-based Analytics
- ML Concepts
- Scripting, reporting & Data Visualization

### Data Analysts

Skills &  
Job roles

- Adobe & Google Analytics
- Programming knowledge
- Scripting & Statistical skills
- Reporting & Data Visualization
- SQL / database

Job Roles	Data Analyst	Data Engineer	Data Scientist
<ul style="list-style-type: none"> <li>→ Pre-Processing &amp; data gathering</li> <li>→ Emphasis on representing data reporting &amp; visualization</li> <li>→ data interpretation</li> <li>→ data acquisition &amp; maintenance</li> </ul>	<ul style="list-style-type: none"> <li>→ Develop, test &amp; maintain architectures</li> <li>→ Understand programming</li> <li>→ Deploy ML &amp; Statistical models</li> <li>→ Pipelining data &amp; ETL operation</li> </ul>	<ul style="list-style-type: none"> <li>→ Responsible for developing Operational models</li> <li>→ Carry out data Analytics using ML &amp; DL</li> <li>→ Strategic Planning</li> <li>→ fill gap b/w Stakeholders &amp; customers</li> </ul>	

## → Datafication

Transformation of Social actions into online Quantified data.. follows:-

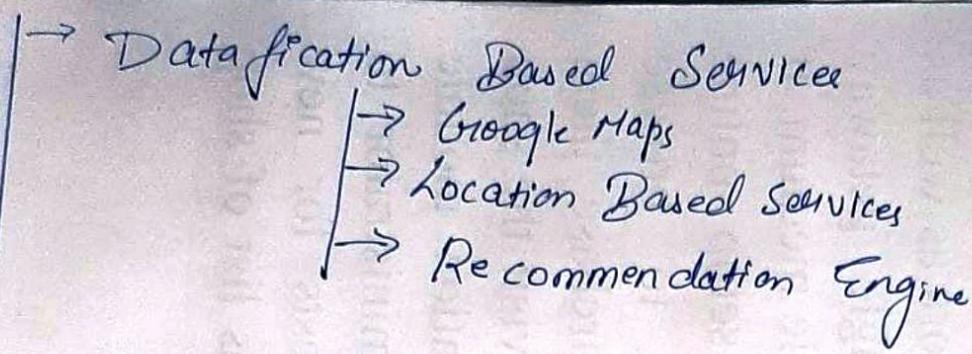
- Real time tracking
- Data Monitoring
- Predictive analysis & optimization

## → Data Driven Organization

(Collective tools, techniques & processes used to transform an organization to a data-driven enterprise.)

## → Social Media Datafication

- |                         |                           |
|-------------------------|---------------------------|
| → Locations Visited     | → Types of media Viewed   |
| → Types of devices Used | → Interaction with People |



## What Is Data Science ?

Data Science is one of the most trending & End demanding technologies right now. It's all about deriving hidden insights from the data in order to solve real world complex problems.

### Process of data Science

- Data Collection
- → Data Cleaning & formatting
- Exploratory data Analysis (EDA)
- Feature Engineering
- Model Building / development Using ML, DL
- Model Evaluation & Validation
- Interpretation & Visualization

## Need of Data Science

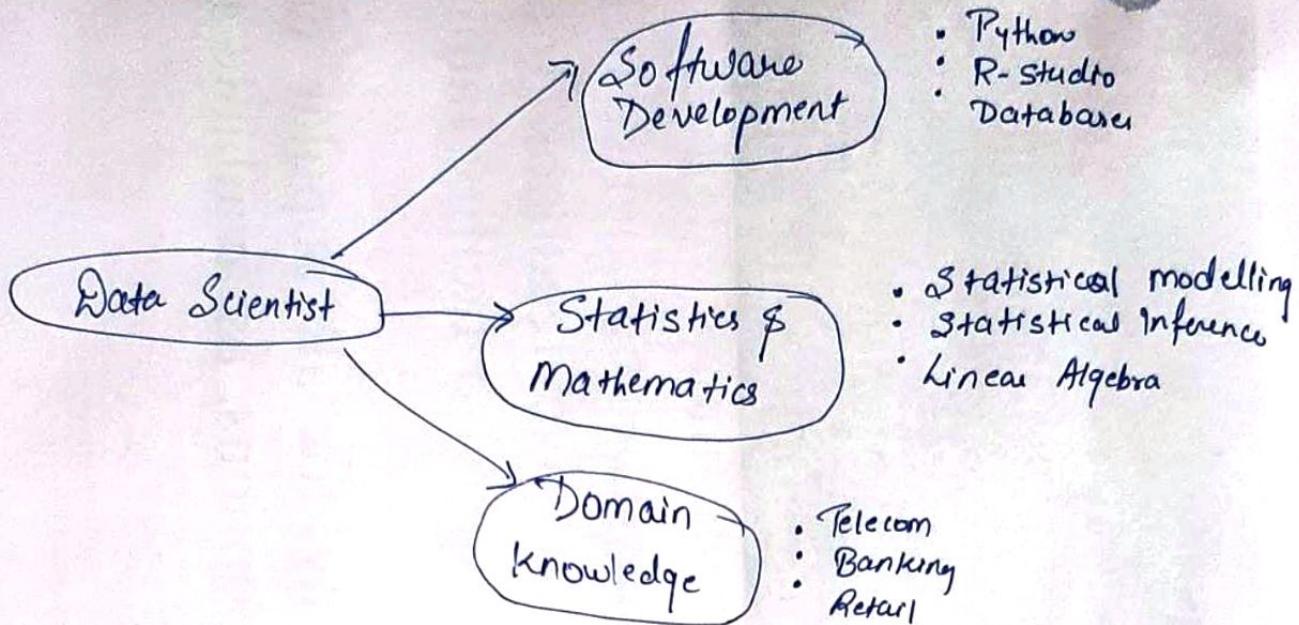
- Handling Big Data
- Informed Decision-making
- Predictive Analytics
- Efficiency & Automation
- Risk Management
- Healthcare Improvement
- Optimizing Processes
- Fraud detection & Security
- Customer Segmentation
- Continuous Improvement
- Scientific Research
- Economic Growth

→ Adapting to change

## What does data Scientist do?

- Model Evaluation & Validation
- Interpretation & Insights
- Data Visualization
- Continuous learning & Improvement
- Deployment & Integration
- Monitoring & Maintenance

## Data Science - Skill set



## Organization - skill set

- Statistics & Mathematics
  - Software Programming
  - Data Wrangling
  - Database Management
  - Data Visualization
  - Machine Learning
  - Cloud Computing
  - Microsoft Excel
  - DevOps
- ⇒ Data Growing Challenge
- ⇒ Infrastructure Improvement Drive

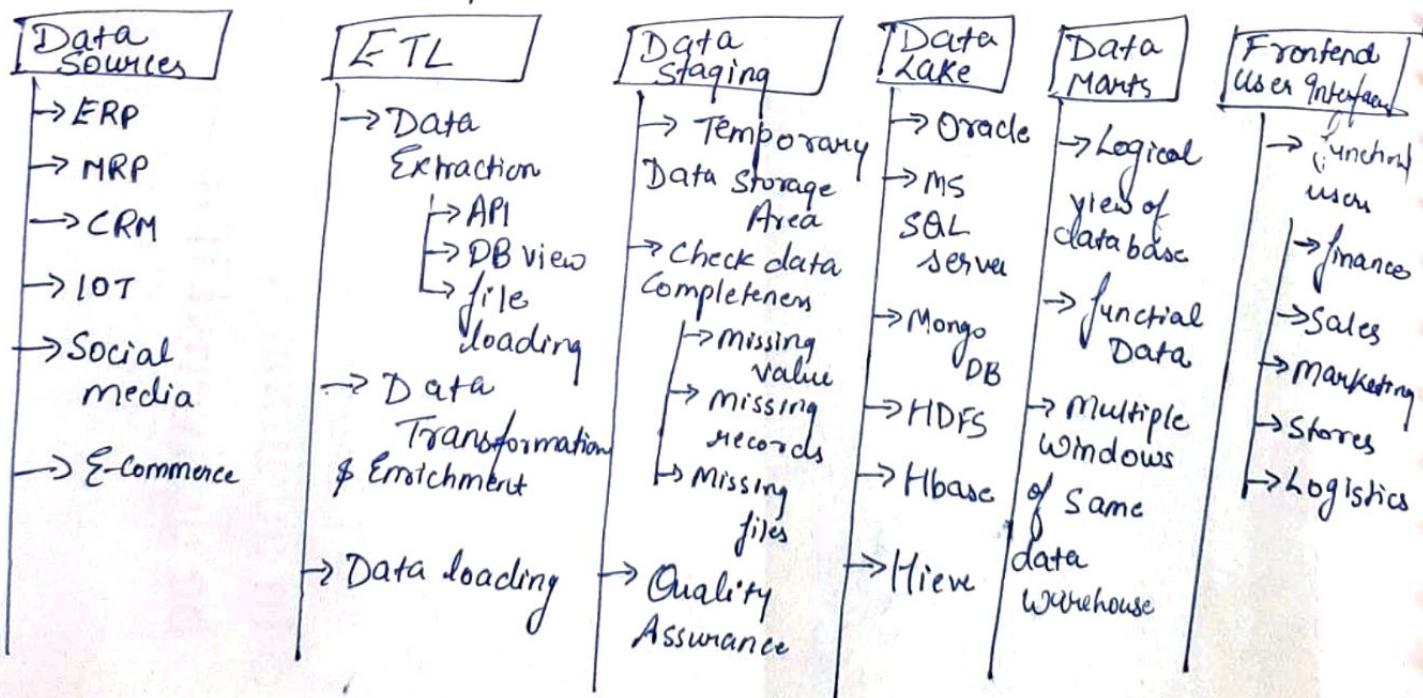
# Expertise

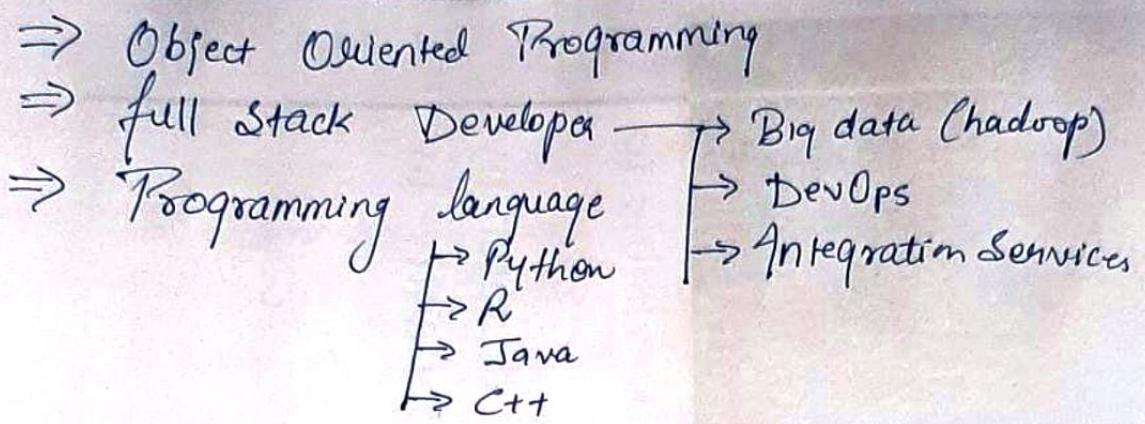
- Statistics
- Domain Knowledge
- Data Engineering
- Data Visualization
- Advance Computing

# Current Landscape

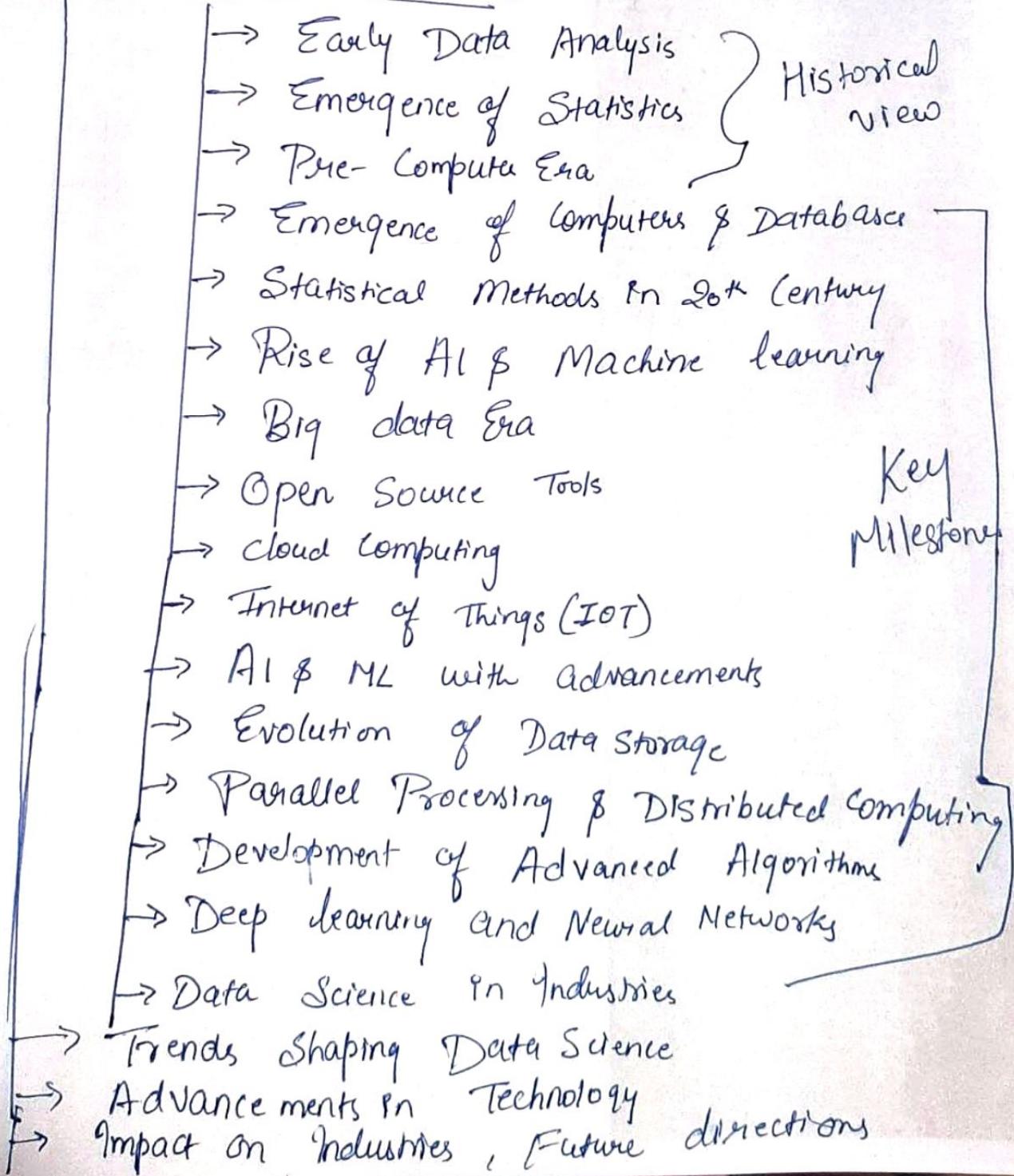
- Discipline → Analysis, Statistics, Machine learning, Domain learning, Awareness
- Focus → Ask Better Questions, Give Better answers, Decision Making At Business scale, Problem
- Discipline → Exploratory, Descriptive, Inferential, Supervised, Targeted & Unsupervised approach
- Tools → Data Analytics, Statistical modeling, ML Programming, S/w def deng.
- Process → Vizulization, Business Process Understanding, Data Collection, Data Preparation, Evaluation, Deployment

# ⇒ Software Development





## Evolution of Data Science



## Stages in data Science Project

- Business Understanding
- Data Mining
- Data Cleaning
- Data Exploration
- Feature Engineering
- Predictive Modeling
- Data Visualization

## Application of Data Science in Various Fields.

- Image Recognition
- Speech Recognition
- Internet Search
- Digital Advertisements
- Recommendation Systems
- Price Comparison Websites
- Banking
- Delivery logistics
- Fraud & Risk Detection  
(And many more)

## Data Security Issue

Data Security is a critical concern in today's digital age, as organizations & individuals increasingly rely on technology to store, process & transmit sensitive information.

- Unauthorized Access
- Data Breaches
- Malware & Ransomware
- Phishing Attacks
- Insider Threats
- Insecure data storage
- Inadequate data Encryption
- Poor Patch Management
- Physical Security Threats
- Data Residency & Compliance
- Insecure APIs
- Lack of User Awareness
- Data Interception

# Data Science for cloud Computing

→ What is cloud?

(On-demand availability of Computer System, resource, especially.)

- Data Storage
- Computing Power
- RAM
- Internet Connectivity

→ Types of Cloud

- Private cloud
- Public cloud
- Hybrid cloud

→ Cloud Service Models

- Software-as-a-Service (SaaS)
- Data-as-a-Service (DaaS)
- Platform-as-a-Service (PaaS)
- Infrastructure-as-a-Service (IaaS)

→ Cloud Infrastructure

→ Data Science with Cloud

(Data Scientist analyses different types of data that are stored in the cloud.)

## Importance of Data Science with cloud

- Storage data in cloud
  - Economical
  - Secure
- Cloud Computing Benefits for data Scientist
  - Open Sources
  - Commercial (Oracle, Tableau)

## Data Science with Cloud: Google Cloud Platform

- What is Google Cloud?
  - Google Search
  - Gmail
  - Google Drive
  - YouTube
  - Google Translate
- Google Cloud Benefits
  - High Productivity & Innovation
  - Easy Adoptability
  - Remote access
  - Security
  - Flexibility & Control
  - Allows to Store & Compute Data
  - Manage SDLC
  - API's Available

## Data Science with cloud: Amazon Web Services

- What is Aws cloud Platform?
  - free trials + 100+ products
  - Compute, Storage, Databases
  - Management Console
  - Machine learning

### AWS cloud Platform Benefits

- Servers
- Storage
- Security
- Networking
- Remote Computing
- Mobile development

## Data Science with cloud: Microsoft Azure cloud

- Hybrid + multi-cloud
- Break, Prepare & Optimize
- Compute, Containers, Networking

### Microsoft Azure Cloud Platform Benefits

- Cost Optimization
- Unmatched Speed of Services
- Global Scale Digital Workplace
- Productivity & Performance
- Flexibility, Reliability & Security
- Azure APIs & Workflows

## Data Science with cloud : Social Media Analytics on Cloud.

- Social Media
- Mobile Appl'
- Data Analysis
- Cloud Computing

→ Social Media Analytics on cloud offers following benefits

- Integrated Data Storage
- Cost Effective
- Robust & fast computation
- Multitasking
- Statistical Modelling
- AI / ML Algorithms